

과적응 감소를 위한 주성분 분석 및 독립성분 분석을 이용한 MLLR 화자적응 알고리즘 개선

Improvement of MLLR Speaker Adaptation Algorithm to Reduce Over-adaptation Using ICA and PCA

김 지 운*, 정 재 호*
(Ji-Un Kim*, Jae-Ho Chung*)

*인하대학교 전자공학과

(접수일자: 2003년 7월 28일; 수정일자: 2003년 9월 18일; 채택일자: 2003년 9월 22일)

본 논문은 MLLR (Maximum Likelihood Linear Regression) 화자 적응시 과적응 방지를 위해 트리 구조에서 HMM 파라메타의 변환을 결정하는 점유 문턱값 (occupation threshold)의 영향을 감소하는 방법에 대해 기술한다. 데이터의 특징을 잘 나타내는 주성분 분석과 독립성분 분석을 통해 모델 혼합성분의 상관관계를 줄이고 상대적으로 데이터의 분포가 적은 축을 삭제함으로써 적은 적응데이터에 의한 과적응의 영향을 감소시켰다. 점유 문턱값을 작게 설정함으로써 변환 함수의 수를 증가시켰을 경우, 기존의 MLLR 알고리즘은 과적응에 의해 화자 독립 모델보다 낮은 인식률을 나타내는 반면, 제안한 MLLR 알고리즘은 화자 독립 모델의 성능에 비해 평균 2% 이상 인식률 향상을 나타내었다.

핵심용어: 화자적응, MLLR, 주성분 분석, 독립성분 분석, 점유 문턱값

투고분야: 음성처리 분야 (2,6)

This paper describes how to reduce the effect of an occupation threshold by that the transform of mixture components of HMM parameters is controlled in hierarchical tree structure to prevent from over-adaptation. To reduce correlations between data elements and to remove elements with less variance, we employ PCA (principal component analysis) and ICA (independent component analysis) that would give as good a representation as possible, and decline the effect of over-adaptation. When we set lower occupation threshold and increase the number of transformation function, ordinary MLLR adaptation algorithm represents lower recognition rate than SI models, whereas the proposed MLLR adaptation algorithm represents the improvement of over 2% for the word recognition rate as compared to performance of SI models.

Keywords: Speaker adaptation, MLLR, PCA, ICA, Occupation threshold

ASK subject classification: Speech signal Processing (2,6)

I. 서론

음성인식 시스템은 화자의 특성의 변화, 음향환경의 변화, 혹은 잘못된 모델링과 같은 훈련환경과 인식환경의 차이에 매우 민감하다. 이러한 환경의 불일치를 보상하는 방법에는 강인하고 환경에 따른 변화가 적은 음성 특징 추출, 모델링 방법의 향상, 적응이나 보상 방법을 이용한 인식 파라미터나 특징 벡터의 수정, 강인한 결정

방법 등이 있다. 특히 적응 데이터를 이용하여 기존의 모델로부터 새로운 음향학적 모델을 유도해 내기 위한 적응 방법은 크게 2가지로 분류된다. 그 중 하나는 변환에 기초한 적응방법으로서 HMM 파라메타를 변환 함수에 의해 변환시키는 방법이다. MLLR과 SM (stochastic matching) 등이 이 부류에 해당한다[1-3]. 이런 방법들은 ML (Maximum Likelihood)을 이용하여 변환 파라메타를 구한다. 또 다른 분류는 MAP (Maximum A Posteriori) 적응 방법에 기인한 여러가지 방법이다. 이는 사전 통계로서 SI HMM 파라메타를 제공함으로써 평균, 공분산, 혼합 성분 가중치 등이 MAP 방법에 의해 적응

책임저자: 김지운 (g2001118@inhavision.inha.ac.kr)
412-751 인천광역시 남구 용현동 253
인하대학교 전자공학과 디지털신호처리실
(전화: 032-860-7420; 팩스: 032-868-3654)

된다[4]. 일반적으로 적응 데이터가 제한된다면 변환에 기반한 적응이 클래스에 의존적인 변환 함수에 의해 모든 HMM 파라미터를 효과적으로 변환시킬 수 있다. 반면 만일 충분한 데이터가 제공되면 MAP방법이 SI HMM 파라미터를 변환함수에 기반한 적응방법보다 효과적으로 적응시킬 수 있다[5,6].

MLLR 화자적응방법은 HMM의 각 혼합 성분들을 몇 개의 클래스로 군집화시킨 2진 회귀나무를 형성하고, 적응 데이터를 각 기저 클래스에 할당한다. 만일 적응 데이터가 적게 할당된 기저 클래스의 혼합 성분들에 대해 점유 확률이 미리 정해진 점유 문턱값보다 작으면 상위 회귀 클래스에서 다시 점유 문턱값과 비교해 혼합 성분의 적응을 결정한다. 만일 점유 문턱값을 크게 설정할 경우, 충분한 적응 데이터가 제공되어도 하위 클래스들의 일반화된 적응 변환 행렬이 추정되므로 적응 속도가 느리고, 점유 문턱값을 작게 설정할 경우, 기저 클래스에서 변이가 큰 혼합 성분들의 적응 변환 행렬을 구할 때 충분한 적응 데이터가 제공되지 않으면 적응 데이터에 편중된 적응 변환 행렬이 추정되므로 인식성능이 나빠진다. 이는 적응 변환 함수를 추정할 때, 혼합 성분들의 분포가 적게 형성된 축에 대한 적응 파라미터들이 적응 데이터에 지나치게 편중되기 때문이다. 즉 점유 문턱값을 결정하는 방법에 따라 화자 적응 속도와 인식율은 큰 변화를 나타낸다. 본 논문은 주성분 분석(PCA: Principle Component Analysis)과 독립성분 분석(ICA: Independent Component Analysis)을 이용하여 문턱값의 변화에 영향을 적게 받도록 MLLR 적응방법을 개선하였다.

II. 주성분 분석 및 독립성분 분석을 이용한 MLLR 알고리즘 개선

MLLR 화자적응방법은 그림 1과 같이 HMM의 각 혼합 성분들을 몇 개의 클래스로 군집화시킨 2진 회귀 나무를 형성하고, 적응 데이터를 각 기저 클래스에 할당한다. 만일 적응 데이터가 적게 할당된 기저 클래스의 혼합 성분들에 대해 점유 확률이 미리 정해진 점유 문턱값보다 작으면 상위 회귀 클래스에서 다시 점유 문턱값과 비교해 혼합 성분의 적응을 결정한다. 만일 점유 문턱값을 크게 설정할 경우, 충분한 적응 데이터가 제공되어도 하위 클래스들의 일반화된 적응 변환 행렬이 추정되므로 적응 속도가 느려지고, 점유 문턱값을 작게 설정할 경우, 기저

클래스 내에서 변이가 큰 혼합 성분들의 적응 변환 행렬을 구할 때 충분한 적응 데이터가 제공되지 않으면 적응 데이터에 편중된 적응 변환 행렬이 추정되므로 인식성능이 나빠진다. 이는 적응 변환 함수를 추정할 때 혼합 성분들의 분포가 적게 형성된 축에 대한 적응 파라미터들이 적응 데이터에 지나치게 편중되기 때문이다. 따라서 본 논문에서는 혼합 성분들의 분포가 적게 형성된 축에 대한 적응 파라미터를 생략함으로써 추정해야 하는 적응 파라미터의 수를 감소시키고 적은 양의 적응데이터에 의한 과적응의 영향을 감소시켰다. 따라서 과적응 방지를 위한 문턱값은 더욱 낮게 설정될 수 있고, 고정된 점유 문턱값에 대해 화자 적응 속도를 향상시킬 수 있다.

2.1. 개선된 MLLR 알고리즘

MLLR 알고리즘의 목적은 각 혼합 성분의 평균 벡터의 변환을 추정하여, SI모델을 새로운 화자의 SD모델로 바꾸는 것이다. 평균 μ_s 를 갖는 혼합 성분 s 에 대해 적응된 평균 $\hat{\mu}_s$ 는 다음과 같이 나타난다.

$$\hat{\mu}_s = P^{-1}W_s P \xi_s \tag{1}$$

여기서 W_s 는 차수 $m \times (m+1)$ 인 변환 행렬이고 P 는 차수 $(m+1) \times (n+1)$ 인 차수감소 행렬이다. (단, $m < n$) P^{-1} 은 차수 확장 행렬로서 차수 감소된 벡터를 원래의 차수로 변환하는 행렬이다. ξ_s 는 평균벡터를 확장시킨 것으로

$$\xi_s = \{w, \mu_{s1}, \mu_{s2}, \dots, \mu_{sn}\} \tag{2}$$

여기서 w 는 오프셋항목이다.

적응데이터를 $O = \{o_1, o_2, \dots, o_T\}$ 라고 가정하자. 여기서 T 는 관측 벡터의 개수이다. 모델 파라미터를 재추정하기 위한 보조 함수는 다음과 같이 나타낼 수 있다.

$$Q(\lambda, \bar{\lambda}) = a - \frac{1}{2} F(O|\lambda) \sum_{t=1}^T \sum_{j=1}^J \gamma_j(t) [n \log(2\pi) + \log |\Sigma_j| + h(o_t, j)] \tag{3}$$

$$h(o_t, j) = (o_t - P^{-1}W_j P \xi_j)' \Sigma_j^{-1} (o_t - P^{-1}W_j P \xi_j) \tag{4}$$

$Q(\lambda, \bar{\lambda})$ 를 $P^{-1}W_j$ 에 대해 일차미분을 행하면 (5)와 같다.

$$\frac{d}{dW_j} Q(\lambda, \bar{\lambda}) = F(O|\lambda) \sum_{t=1}^T \gamma_j(t) \Sigma_j^{-1} [o_t - P^{-1}W_j P \xi_j] \xi_j' P' \tag{5}$$

(5)의 최대값을 구하면 (6)과 같다

$$\sum_{t=1}^T \gamma_s(t) \Sigma_s^{-1} o_t \xi_s^T P^T = \sum_{t=1}^T \gamma_s(t) \Sigma_s^{-1} A_s P \xi_s \xi_s^T P^T \quad (6)$$

여기서 $A_s = P^{-1}W_s$ 이다. 식 (6)의 해를 구하는 자세한 방법은 C. J. Leggetter의 논문을 참고하라[1]. 따라서 과적응에 의한 영향은 평균 벡터를 주성분 혹은 독립성분으로 변환하기 위한 변환 행렬 P 에 의존한다.

2.2. 변환행렬 P 의 추정

본 논문에서는 추정해야 하는 적응 파라메타 수의 감소를 위해 주성분 분석 및 독립성분 분석을 적용한다. 주성분 분석과 독립성분 분석을 적용하기 전에 먼저 SI모델의 평균값을 몇 개의 클래스로 그룹화하는 작업이 필요한데 본 논문에서는 기존의 MLLR 방법을 수행하기 전에 만들어진 M. J. F. Gales의 회귀나무[7]를 이용한다. 회귀나무를 구성하는 방법은 다음과 같다.

- ① 분할될 터미널 노드를 선택한다.
- ② 선택된 노드에 속해 있는 혼합 성분들의 평균과 분산을 구한다.
- ③ 부모 노드의 평균을 기준으로 분산에 비례하여 반대 방향의 두 점을 자식 노드의 초기 평균으로 정한다.
- ④ 부모 노드에 속해 있는 혼합 성분들을 Euclidean 거리를 이용하여 각 자식 노드에 나눈다.
- ⑤ 각 자식 노드에 할당된 혼합 성분들을 이용하여 자식 노드의 평균을 다시 구한다.
- ⑥ 평균 값의 변화가 없을 때까지 ④번을 반복한다.

회귀 나무를 형성한 후, 그림 1과 같이 회귀 나무의 기저 클래스의 각 평균 벡터들을 연결하여 하나의 슈퍼 벡터, s ,로 만든 후 이들을 하나의 열벡터로 하는 행렬, $S = [s_1, s_2, \dots, s_k]$ 를 형성한다. 여기서 k 는 회귀 나무의 기저 클래스의 개수이다. 이와 같이 형성된 행렬 S 에 대해 주성분 분석 또는 독립성분 분석을 행한다.

주성분 분석의 목적은 데이터 성분들간의 상관관계를 줄이고 데이터의 특성을 잘 나타내면서 적은 차수를 갖는 변수를 찾아내는데 있다. 즉, 데이터의 분산을 최대로 하는 축을 찾아서 축 변환에 의해서 각 데이터들의 상관성을 최대로 줄임으로써 데이터를 새로운 하위영역으로 사상하는 것이기 때문에 각 데이터의 차원을 줄일 수 있다. 본 논문에서 적용한 주요성분 분석 방법의 순서는 다음과

Regression tree

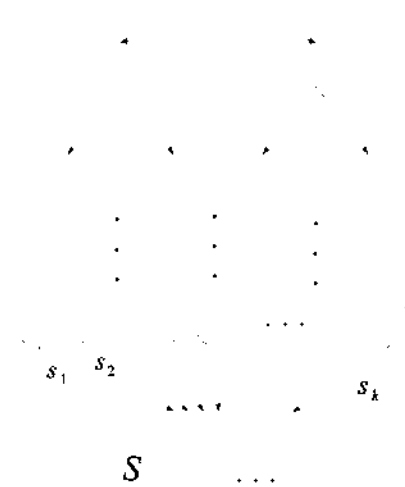


그림 1. 슈퍼벡터 행렬 형성
Fig. 1. Constructing a supervector matrix S .

같다.

- ① S^T 의 공분산 행렬 C' 를 구한다.

$$C' = E\{(S^T - E(S^T))(S^T - E(S^T))^T\} = E\{(S - E(S))^T(S - E(S))\} = \frac{1}{k} \psi^T \psi \quad (7)$$

- ② C' ($=\psi^T \psi$)의 고유 벡터 행렬 W 와 고유값 대각 행렬 Λ 를 구한다.

$$\psi^T \psi V = \Lambda W \quad (8)$$

- ③ 양변에 ψ 를 곱한다.

$$\psi \psi^T (\psi W) = \Lambda (\psi W) \quad (9)$$

- ④ S 의 공분산 행렬 C 의 고유벡터 행렬과 고유값 행렬로써 ψW 와 Λ 를 사용한다.
- ⑤ 고유벡터 행렬 $V = \psi W$ 를 이용하여 혼합성분의 주 성분을 구한다.

$$y = V\mu \quad (10)$$

주성분 분석을 통해 얻어진 고유 벡터 행렬 $V = [v_1, v_2, \dots, v_k]$ 을 이용하여 각 혼합 성분의 평균 벡터를 고유 벡터 영역으로 사상시킬 수 있다. 이때 고유값에 의거하여 사용하는 고유벡터의 수를 조절할 수 있다. 작은

고유값을 갖는 고유벡터는 0에 가깝기 때문에, 작은 고유값은 생략함으로써 HMM의 평균 파라메타의 차원이 줄어들고 추정해야 하는 적응 파라메타의 수가 감소한다. 즉, 변환된 데이터의 분포의 변별력이 상대적으로 작은 고유벡터는 사용하지 않음으로써 혼합 성분의 평균 벡터의 차수를 줄일 수가 있다. 따라서 m개의 고유 벡터를 사용한다면 변환행렬 P는 다음과 같이 나타낼 수 있다.

$$P = \begin{bmatrix} 1 & 0 \cdots 0 \\ 0 & \\ \vdots & V' \\ 0 & \end{bmatrix} \quad (11)$$

여기서 $V' = [v_1, v_2, \dots, v_m]$ 이다.

독립성분 분석의 목적은 주성분 분석의 목적과 유사하다. 그러나 주성분 분석에서 잉여 정보량은 데이터의 각 성분들 사이의 상관관계에 의해 추정되는 반면, 독립성분 분석에서는 각 성분의 독립성의 개념이 더 추가된다. 즉, 독립성분 분석은 출력신호가 통계학적으로 독립인 선형 변환 시스템을 구하는 것이다. 본 논문에서 사용한 독립성분 분석 방법은 A. Hyvarinen의 FastICA로써 다음과 같다[8].

- ① S의 화이트닝 변환을 수행한다.

$$Z = \Lambda^{-1/2} W^T S \quad (12)$$

- ② 추정할 독립성분의 개수 m을 결정하고, 카운터 p를 1로 초기화한다.
- ③ 가중 행렬 A의 열벡터 a_p 의 놈 (norm)을 초기화한다.
- ④ 식 (13)을 이용하여 열벡터 a_p 를 갱신한다.

$$a_p = E\{z_i g(a_p^T z_i)\} - E\{g'(a_p^T z_i)\} a_p \quad (13)$$

여기서, $g(y) = y^3$ 이다.

- ⑤ 열벡터 a_p 를 직교화한다.

$$a_p = a_p - \sum_{j=1}^{p-1} (a_p^T a_j) a_j \quad (14)$$

- ⑥ 열벡터 a_p 를 정규화한다 ($a_p = a_p / \|a_p\|$).
- ⑦ 열벡터 a_p 가 수렴할 때까지 ④~⑥을 반복한다.
- ⑧ p를 하나 증가시키고, p가 m이 될 때까지 ③~⑦을 반복한다.

주성분 분석과 같은 신호의 상관관계에 기반한 변환 방법과 비교해 볼 때, 독립성분 분석은 입력신호의 상관관계 (2nd order statistics)뿐만 아니라 더 높은 차원의 통계학적인 의존관계를 줄인다. 따라서 HMM 파라메타의 차원을 줄일 때에는 독립성분 분석이 주성분 분석보다 우수한 성능을 보일 수 있다. 그러나 입력 데이터인 HMM 파라메타에 독립성분이 충분히 포함되어 있지 않다면, 주성분 분석이 독립성분분석보다 데이터의 특징을 잘 나타낼 수 있다[7]. 주성분 분석에서와 유사한 방법으로 독립성분 분석에 의한 가중행렬을 이용해서 변환행렬 P를 구할 수 있다.

III. 실험 및 실험 결과

본 논문에서 사용한 데이터 베이스는 PBW (Phonetic Balanced Word) DB이다. 31명의 화자가 발성한 445개의 단어를 훈련 데이터로 사용하였다. 사용한 음성의 특징 파라메타는 18차의 LPCC (Linear Predictive Cepstral Coefficient)와 델타 계수를 사용하였고, CMS (Cepstral Mean Subtraction)을 행하였다. 음성인식은 40개의 단음 소로 훈련된 모델을 이용하여 행하였다. 각 단음소 모델은 3개의 상태를 가지고 있고 각 상태는 8개의 혼합 성분을 포함한다. MLLR 방법과 제안한 알고리즘의 실험에서는 블럭의 개수가 2개인 블럭 대각 변환 행렬을 사용하였다. 단 제안한 알고리즘에서는 각각의 블럭을 따로 독립성분 분석과 주성분 분석을 행하였다. PCA와 ICA를 적용할 때, 독립성분과 주성분의 차수는 MLLR 알고리즘만을 적용하였을 때와 유사한 성능을 나타내는 10차를 사용하였다. 그러나 차수의 변화에 따른 점유 문턱값이 인식 성능에 미치는 영향은 대동소이한 것으로 나타났다.

표 1에 MLLR 화자적응만을 적용하였을 때 점유 문턱값에 따른 인식 결과를 나타내었다. 많은 양의 적응데이터가 제공될 경우 (3sets) 점유 문턱값에 따른 인식율의 최대값과 최소값의 차이는 약 1%이며, 인식율의 분산은 0.19로 점유 문턱값의 영향을 적게 받는 것으로 나타났다. 반면 적은 양의 적응 데이터가 제공될 경우 (1set) 인식율의 최대값과 최소값의 차이는 약 8.5%이고, 분산은 7.9로 나타났다. 특히 점유 문턱값이 작아질 경우 SI 모델의 인식율보다 낮은 인식 성능을 나타내었다. 이는 적은 적응 데이터가 제공될 경우 과적응에 기인한 것으로 사료된다.

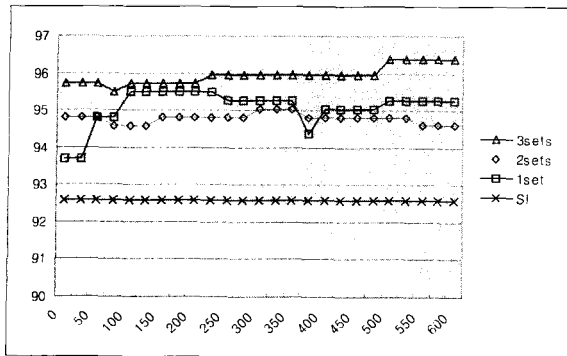


그림 4. 독립성분 분석을 적용한 MLLR 화자적응에서 점유 문턱값에 따른 인식율 변화
 Fig. 4. Recognition rate changes by occupation threshold in MLLR with ICA.

해 점유 문턱값에 많은 영향을 받는 것으로 나타났다. 그러나 MLLR 알고리즘만을 적용하였을 때 과적응에 의해 SI모델의 인식성능보다 낮은 인식율을 보였던 점유 문턱값에 대해서 주성분 분석이나 독립성분 분석을 적용하였을 때 최대 6%이상 인식율이 향상되었다. 특히 독립성분 분석을 적용하였을 때 점유 문턱값에 영향을 적게 받는 것으로 나타났는데, 이는 독립성분 분석이 통계학적으로 2차원적인 분석방법인 신호의 상관관계를 줄였을 뿐 아니라, 더 높은 차원의 통계적 의존관계를 줄였기 때문으로 사료된다.

IV. 결론 및 향후 과제

본 논문은 MLLR의 연산량 감소를 위해 독립성분 분석과 주성분 분석을 이용하여 적응데이터가 충분히 분포되지 않은 축을 삭제함으로써 과적응 방지를 위한 점유 문턱값의 영향을 감소시켰다. 적응 데이터가 충분히 제공되지 않을 경우 기존의 MLLR 알고리즘은 점유 문턱값에 따라 인식율의 변화가 심하게 나타났고, 주성분 분석이나 독립성분 분석을 적용하였을 경우 점유 문턱값에 의한 인식율의 변화는 매우 적게 나타났다. 따라서 낮은 점유 문턱값에 의한 많은 적응 변환 함수를 사용할 경우, 적은 양의 적응 데이터에 대한 과적응 영향을 감소시킬 수 있었다.

제안한 알고리즘은 기존의 MLLR 알고리즘에 비해 더 많은 적응 변환 함수를 사용할 수 있으므로 더 깊은 회귀나무를 형성할 수 있다. 따라서 제안한 알고리즘에 적합

한 회귀나무를 형성하는 방법이 요구된다.

참고 문헌

1. C. J. Leggetter, *Improved Acoustic Modelling for HMMs Using Linear Transforms*, PhD Thesis, Univ. of Cambridge Feg. 1995.
2. A. Sankar and C. H. Lee, "A maximum-likelihood approach to stochastic matching for robust speech recognition," *IEEE Trans. on Speech Audio Processing*, 4, 190-202, 1996.
3. V. Digalakis, "On-line adaptaion of hidden Markov models using incremental estimation algorithms," *Proc. 5th Eur. Conf. Speech Communication and Technology*, 4, 1859-1862, 1997.
4. C. H. Lee, C. H. Lin, and B. H. Juang, "A study on speaker adaptation of the parameters of continuous density hidden Markov models," *IEEE Trans. on Signal Processing*, 39 (4), 806-814, April 1991.
5. O. Siohan, C. Chesla, and C. H. Lee, "Joint maximum a posteriori adaptation of transformation and HMM parameters," *IEEE Trans. on Speech and Audio Processing*, 9 (4), 417-428, May 2001.
6. J. T. Chien, "Online hierarchical transformation of hidden Markov models for speech recognition," *IEEE Trans. on Speech and Audio Processing*, 7 (6), 656-667, Nov. 1999.
7. M. J. F. Gales, "Maximum likelihood linear tranformation for HMM-based speech recognition," Cambridge Univ. Eng. Dept. Tr 291, May 1977.
8. A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Willy-Interscience, 147-164, 2001.
9. M. E. Tipping, and C. M. Bishop, "Probabilistic Principal Component Analysis," *Journal of the Royal Statistical Society, Series B*, 61, Part 3, 611-612, 1999.

저자 약력

● 김 지 운 (Ji-Un Kim)



1998년2월: 인하대학교 전자공학과 공학사
 2000년2월: 인하대학교 전자공학과 공학석사
 2000년3월~ 현재: 인하대학교 전자공학과 박사과정
 * 주관심분야: 음성인식, 발화인중, 화자인식, 화자 적응

● 정 재 호 (Jae-Ho Chung)

1982년: University of Maryland (BSEE)
 1984년: University of Maryland (MSEE)
 1990년: Georgia Institute of Technology (Ph. D.)
 1984년~1985년: 미국 국방성 산하 해군 연구소, 신호처리실, 연구원
 1991년~1992년: AT&T Bell Laboratories, 음성신호처리연구실, 연구원 (MTS)
 1992년~ 현재: 인하대학교 공과대학 전자공학과, 정교수